(19)日本国特許庁（ＪＰ）　(12)公開特許公報（Ａ）　(11)特許出願公開番号

特開平9－319394

(43)公開日　平成９年(1997)12月12日

(51)Int.Cl.[6]	識別記号	庁内整理番号	FI		技術表示箇所
G10L 5/04			G10L 5/04	E	
				A	
				F	
3/00			3/00	H	

審査請求　未請求　請求項の数９　ＯＬ　（全 17 頁）

(21)出願番号　特願平9－46694

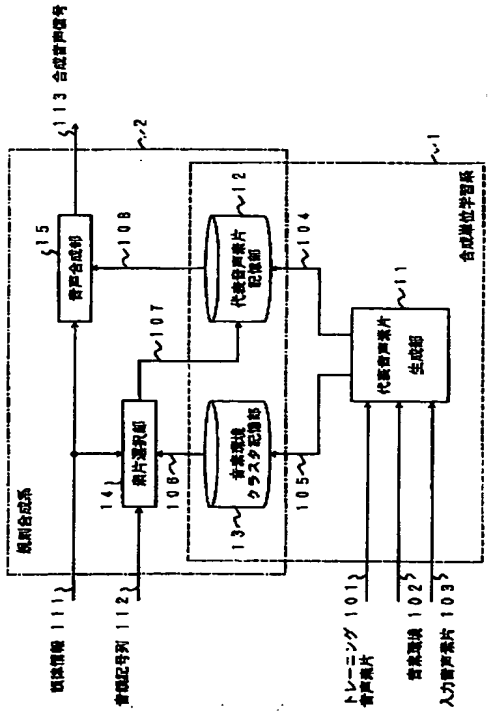
(22)出願日　平成９年(1997)２月28日

(31)優先権主張番号　特願平8－54714
(32)優先日　平８(1996)３月12日
(33)優先権主張国　日本（ＪＰ）
(31)優先権主張番号　特願平8－77393
(32)優先日　平８(1996)３月29日
(33)優先権主張国　日本（ＪＰ）

(71)出願人　000003078
株式会社東芝
神奈川県川崎市幸区堀川町72番地
(72)発明者　篭嶋　岳彦
神奈川県川崎市幸区小向東芝町１番地　株式会社東芝研究開発センター内
(72)発明者　赤嶺　政巳
神奈川県川崎市幸区小向東芝町１番地　株式会社東芝研究開発センター内
(74)代理人　弁理士　鈴江　武彦　（外６名）

(54)【発明の名称】　音声合成方法

(57)【要約】
【課題】テキスト音声合成による合成音声の音質を効果的に向上させることができる音声合成方法を提供する。
【解決手段】代表音声素片生成部１１において音素環境１０２がラベル付けされたトレーニング音声素片１０１のピッチ・継続時間長に従って入力音声素片１０３のピッチ・継続時間長を変更して複数の合成音声素片を生成し、合成音声素片とトレーニング音声素片１０１との間の距離尺度に基づいて入力音声素片１０３から代表音声素片１０４を選択して代表音声素片記憶部１２に記憶し、さらに距離尺度に基づいて代表音声素片にそれぞれ対応する複数の音素環境クラスタ１０５を生成して音素環境クラスタ記憶部１３に記憶し、代表音声素片記憶部１２から入力音素の音素環境を含む音素環境クラスタに対応する代表音声素片を読み出して音声合成部１５で接続することにより、合成音声信号１１３を生成する。



1

【特許請求の範囲】

【請求項１】複数の第１の音声素片のピッチおよび継続時間長の少なくとも一方に従って複数の第２の音声素片のピッチおよび継続時間長の少なくとも一方を変更することにより複数の合成音声素片を生成し、
これらの合成音声素片と前記第１の音声素片との間の距離尺度に基づいて前記第２の音声素片から複数の代表音声素片を選択して記憶し、
これらの代表音声素片から所定の代表音声素片を選択して接続することによって音声を合成することを特徴とする音声合成方法。

【請求項２】音素環境がラベル付けされた複数の第１の音声素片のピッチおよび継続時間長の少なくとも一方に従って複数の第２の音声素片のピッチおよび継続時間長の少なくとも一方を変更して複数の合成音声素片を生成し、
これらの合成音声素片と前記第１の音声素片との間の距離尺度に基づいて前記第２の音声素片から複数の代表音声素片を選択して記憶し、
前記距離尺度に基づいて前記代表音声素片にそれぞれ対応する複数の音素環境クラスタを生成し、
前記代表音声素片から入力音素の音素環境を含む音素環境クラスタに対応する代表音声素片を選択して接続することによって音声を合成することを特徴とする音声合成方法。

【請求項３】音素環境がラベル付けされた複数の第１の音声素片のピッチおよび継続時間長の少なくとも一方に従って複数の第２の音声素片のピッチおよび継続時間長の少なくとも一方を変更して複数の合成音声素片を生成し、
これらの合成音声素片と前記第１の音声素片との間の距離尺度に基づいて複数の音素環境クラスタを生成し、
前記距離尺度に基づいて前記第２の音声素片から各音素環境クラスタにそれぞれ対応する複数の代表音声素片を選択して記憶し、
これらの代表音声素片から所定の代表音声素片を選択して接続することによって音声を合成することを特徴とする音声合成方法。

【請求項４】音素環境がラベル付けされた複数の第１の音声素片のピッチおよび継続時間長の少なくとも一方に従って複数の第２の音声素片のピッチおよび継続時間長の少なくとも一方を変更して複数の合成音声素片を生成し、
これらの合成音声素片と前記第１の音声素片との間の距離尺度に基づいて複数の音素環境クラスタを生成し、
前記距離尺度に基づいて前記第２の音声素片から各音素環境クラスタにそれぞれ対応する複数の代表音声素片を選択して記憶し、
これらの代表音声素片から入力音素の音素環境を含む音素環境クラスタに対応する代表音声素片を選択して接続

2

することによって音声を合成することを特徴とする音声合成方法。

【請求項５】複数の第１の音声素片のピッチおよび継続時間長の少なくとも一方に従って複数の第２の音声素片のピッチおよび継続時間長の少なくとも一方を変更することにより複数の合成音声素片を生成し、
これらの合成音声素片についてスペクトル整形を行い、
このスペクトル整形を行った後の各合成音声素片と前記第１の音声素片との間の距離尺度に基づいて前記第２の音声素片から複数の代表音声素片を選択して記憶し、
これらの代表音声素片から所定の代表音声素片を選択して接続することによって音声を合成し、
この合成した音声のスペクトル整形を行って最終的な合成音声を生成することを特徴とする音声合成方法。

【請求項６】音素環境がラベル付けされた複数の第１の音声素片のピッチおよび継続時間長の少なくとも一方に従って複数の第２の音声素片のピッチおよび継続時間長の少なくとも一方を変更して複数の合成音声素片を生成し、
これらの合成音声素片についてスペクトル整形を行い、
このスペクトル整形を行った後の各合成音声素片と前記第１の音声素片との間の距離尺度に基づいて前記第２の音声素片から複数の代表音声素片を選択して記憶し、
前記距離尺度に基づいて前記代表音声素片にそれぞれ対応する複数の音素環境クラスタを生成し、
前記代表音声素片から入力音素の音素環境を含む音素環境クラスタに対応する代表音声素片を選択して接続することによって音声を合成し、
この合成した音声のスペクトル整形を行って最終的な合成音声を生成することを特徴とする音声合成方法。

【請求項７】音素環境がラベル付けされた複数の第１の音声素片のピッチおよび継続時間長の少なくとも一方に従って複数の第２の音声素片のピッチおよび継続時間長の少なくとも一方を変更して複数の合成音声素片を生成し、
これらの合成音声素片についてスペクトル整形を行い、
このスペクトル整形を行った後の各合成音声素片と前記第１の音声素片との間の距離尺度に基づいて複数の音素環境クラスタを生成し、
前記距離尺度に基づいて前記第２の音声素片から各音素環境クラスタにそれぞれ対応する複数の代表音声素片を選択して記憶し、
これらの代表音声素片から所定の代表音声素片を選択して接続することによって音声を合成し、
この合成した音声のスペクトル整形を行って最終的な合成音声を生成することを特徴とする音声合成方法。

【請求項８】音素環境がラベル付けされた複数の第１の音声素片のピッチおよび継続時間長の少なくとも一方に従って複数の第２の音声素片のピッチおよび継続時間長の少なくとも一方を変更して複数の合成音声素片を生成

し、
これらの合成音声素片についてスペクトル整形を行い、
このスペクトル整形を行った後の各合成音声素片と前記第１の音声素片との間の距離尺度に基づいて複数の音素環境クラスタを生成し、
前記距離尺度に基づいて前記第２の音声素片から各音素環境クラスタにそれぞれ対応する複数の代表音声素片を選択して記憶し、
これらの代表音声素片から入力音素の音素環境を含む音素環境クラスタに対応する代表音声素片を選択して接続することによって音声を合成し、
この合成した音声のスペクトル整形を行って最終的な合成音声を生成することを特徴とする音声合成方法。
【請求項９】複数の第１の音声素片のピッチおよび継続時間長の少なくとも一方に従って代表音声素片を用いて複数の合成音声素片を生成し、
これらの合成音声素片と前記複数の第１の音声素片との間で定義される歪みの評価関数に基づいて複数の前記代表音声素片を求めて記憶し、
これらの代表音声素片から所定の代表音声素片を選択して接続することによって音声を合成することを特徴とする音声合成方法。
【発明の詳細な説明】
【０００１】
【発明の属する技術分野】本発明は、テキスト音声合成のための音声合成方法に係り、特に音韻記号列、ピッチおよび音韻継続時間長などの情報から音声信号を生成する音声合成方法に関する。
【０００２】
【従来の技術】任意の文章から人工的に音声信号を作り出すことをテキスト音声合成という。テキスト音声合成は、一般的に言語処理部、音韻処理部および音声合成部の３つの段階によって行われる。入力されたテキストは、まず言語処理部において形態素解析や構文解析などが行われ、次に音韻処理部においてアクセントやイントネーションの処理が行われて、音韻記号列・ピッチ・音韻継続時間長などの情報が出力される。最後に、音声信号合成部で音韻記号列・ピッチ・音韻継続時間長などの情報から音声信号を合成する。そこで、テキスト音声合成に用いる音声合成方法は、任意の音韻記号列を任意の韻律で音声合成することが可能な方法でなければならない。
【０００３】このような任意の音韻記号列を音声合成する音声合成装置の原理は、母音をＶ、子音をＣで表すと、ＣＶ、ＣＶＣ、ＶＣＶといった基本となる小さな単位の特徴パラメータ（これを代表音声素片という）を記憶し、これらを選択的に読み出した後、ピッチや継続時間長を制御して接続することにより、音声を合成するというものである。従って、記憶されている代表音声素片が合成音声の品質を大きく左右することになる。

【０００４】従来、これらの代表音声素片の作成はもっぱら人手に頼っており、音声信号の中から試行錯誤的に切り出してくる場合がほとんどであるため、膨大な労力を要していた。このような代表音声素片作成の作業を自動化し、音声合成に使用するのに適した代表音声素片を容易に生成する方法として、例えば音素環境クラスタリング（ＣＯＣ）と呼ばれる技術が特開昭６４－７８３００「音声合成方法」に開示されている。
【０００５】ＣＯＣの原理は、音素名や音素環境のラベルを多数の音声素片に付与し、そのラベルが付与された音声素片を音声素片間の距離尺度に基づいて音素環境に関する複数のクラスタに分類し、その各クラスタのセントロイドを代表音声素片とするものである。ここで、音素環境とは当該音声素片にとっての環境となる要因全ての組合せであり、その要因としては当該音声素片の音素名、先行音素、後続音素、後々続音素、ピッチ周期、パワー、ストレスの有無、アクセント核からの位置、息継ぎからの時間、発声速度、感情などが考えられる。実音声中の各音素は音素環境によって音韻が変化しているため、音素環境に関する複数のクラスタ毎に代表素片を記憶しておくことにより、音素環境の影響を考慮した自然な音声を合成することが可能となっている。
【０００６】
【発明が解決しようとする課題】上に述べたように、テキスト音声合成のための音声合成では、代表音声素片のピッチや継続時間長を指定された値に変更して合成する必要がある。このようなピッチや継続時間長の変更により、代表音声素片を切り出してきた音声信号の音質と比較して合成音声の音質がある程度劣化することになる。
【０００７】これに対して、上記のＣＯＣによるクラスタリングでは、音声素片間の距離尺度に基づいてクラスタリングを行っているにすぎないため、合成の際のピッチや継続時間の変更の効果が全く考慮されていないという問題がある。すなわち、ＣＯＣによるクラスタリングおよび各クラスタの代表音声素片は、実際にピッチや継続時間長を変更して合成された合成音声のレベルでは、必ずしも適当なものになっているという保証はない。
【０００８】本発明は、このような問題点を解決すべくなされたものであり、テキスト音声合成による合成音声の音質を効果的に向上させることができる音声合成方法を提供することを目的とする。
【０００９】
【課題を解決するための手段】上記の課題を解決するため、本発明はピッチや継続時間長の変更の影響を考慮して、合成音声のレベルで自然音声に対する歪みが小さくなるような代表音声素片を生成し、その代表音声素片を用いて音声を合成することにより、自然音声に近い合成音声を生成するようにしたものである。
【００１０】すなわち、本発明に係る音声合成方法は、複数の第１の音声素片のピッチおよび継続時間長の少な

くとも一方に従って複数の第２の音声素片のピッチおよび継続時間長の少なくとも一方を変更することにより複数の合成音声素片を生成し、これらの合成音声素片と第１の音声素片との間の距離尺度に基づいて第２の音声素片から複数の代表音声素片を選択して記憶し、これらの代表音声素片から所定の代表音声素片を選択して接続することによって音声を合成することを特徴とする。

【００１１】ここで、第１および第２の音声素片は、ＣＶ，ＶＣＶ，ＣＶＣといった音声合成単位で音声信号中から切り出される素片であり、切り出された波形もしくはその波形から何らかの方法で抽出されたパラメータ系列などを表すものとする。これらのうち、第１の音声素片は合成音声の歪みを評価するために用いられ、また第２の音声素片は代表音声素片の候補として用いられる。合成音声素片は、第２の音声素片に対して少なくともピッチまたは継続時間長を変更して生成される合成音声波形またはパラメータ系列などを表す。

【００１２】合成音声素片と第１の音声素片との間の距離尺度によって、合成音声の歪みが表わされる。従って、この距離尺度つまり歪みがより小さくなる音声素片を第２の音声素片から選択して代表音声素片として記憶しておき、これらの代表音声素片から所定の代表音声素片を選択して接続すれば、自然音声に近い高品質の合成音声が生成される。

【００１３】本発明の第１の態様では、音素環境がラベル付けされた複数の第１の音声素片のピッチおよび継続時間長の少なくとも一方に従って複数の第２の音声素片のピッチおよび継続時間長の少なくとも一方を変更して複数の合成音声素片を生成し、これらの合成音声素片と第１の音声素片との間の距離尺度に基づいて第２の音声素片から複数の代表音声素片を選択して記憶し、前記距離尺度に基づいて代表音声素片にそれぞれ対応する複数の音素環境クラスタを生成し、複数の代表音声素片から入力音素の音素環境を含む音素環境クラスタに対応する代表音声素片を選択して接続することによって音声を合成する。

【００１４】ここで、音素環境とは前述した通り音声素片にとっての環境となる要因、例えば当該音声素片の音素名、先行音素、後続音素、後々続音素、ピッチ周期、パワー、ストレスの有無、アクセント核からの位置、息継ぎからの時間、発声速度、感情といった要素の組み合わせであり、音素環境クラスタとは言い換えれば音素環境の集合であり、例えば「当該素片の音韻が／ｋａ／、先行音韻が／ｉ／または／ｕ／、ピッチ周波数が２００Ｈｚ以下」というようなものを意味する。

【００１５】第１の態様のように、距離尺度つまり合成音声の歪みに基づいて代表音声素片にそれぞれ対応する複数の音素環境クラスタを生成し、入力音素の音素環境を含む音素環境クラスタに対応する代表音声素片を選択して接続するようにすれば、例えば同一音素名の音声素片が複数の音素環境に存在する場合でも、実際の入力音素の音素環境が含まれる音素環境クラスタに対応する代表音声素片のみが選択されることにより、より自然な合成音声が得られる。

【００１６】本発明の第２の態様では、音素環境がラベル付けされた複数の第１の音声素片のピッチおよび継続時間長の少なくとも一方に従って複数の第２の音声素片のピッチおよび継続時間長の少なくとも一方を変更して複数の合成音声素片を生成し、これらの合成音声素片と第１の音声素片との間の距離尺度に基づいて複数の音素環境クラスタを生成し、前記距離尺度に基づいて第２の音声素片から各音素環境クラスタにそれぞれ対応する複数の代表音声素片を選択して記憶し、これらの代表音声素片から所定の代表音声素片を選択して接続することによって音声を合成する。この第２の態様は、音声素片が一つの音素環境にのみ存在する場合に有効である。

【００１７】本発明の第３の態様では、音素環境がラベル付けされた複数の第１の音声素片のピッチおよび継続時間長の少なくとも一方に従って複数の第２の音声素片のピッチおよび継続時間長の少なくとも一方を変更して複数の合成音声素片を生成し、これらの合成音声素片と第１の音声素片との間の距離尺度に基づいて複数の音素環境クラスタを生成し、第１の音声素片と合成音声素片との間の距離尺度に基づいて第２の音声素片から各音素環境クラスタにそれぞれ対応する複数の代表音声素片を選択して記憶し、これらの代表音声素片から入力音素の音素環境を含む音素環境クラスタに対応する代表音声素片を選択して接続することによって音声を合成する。

【００１８】この第３の態様によっても、第１の態様と同様に、例えば同一音素名の音声素片が複数の音素環境に存在する場合、実際の入力音素の音素環境が含まれる音素環境クラスタに対応する代表音声素片のみが選択されることにより、より自然な合成音声が得られる。

【００１９】また、本発明に係る他の音声合成方法は、複数の第１の音声素片のピッチおよび継続時間長の少なくとも一方に従って複数の第２の音声素片のピッチおよび継続時間長の少なくとも一方を変更することにより複数の合成音声素片を生成し、さらにこれらの合成音声素片についてスペクトル整形を行い、このスペクトル整形を行った後の各合成音声素片と第１の音声素片との間の距離尺度に基づいて第２の音声素片から複数の代表音声素片を選択して記憶し、これらの代表音声素片から所定の代表音声素片を選択して接続することによって音声を合成し、この合成した音声のスペクトル整形を行って最終的な合成音声を生成することを特徴とする。

【００２０】この場合、先に示した第１、第２および第３の態様においても、複数の合成音声素片を生成した後、スペクトル整形を行うようにする。ここで、スペクトル整形は「めりはり」のある明瞭な音声を合成するための処理であり、例えばホルマント強調やピッチ強調を

行う適応ポストフィルタによるフィルタリングによって実現される。

【００２１】このように代表音声素片の接続によって合成される音声に対してスペクトル整形を行うと共に、合成音声素片に対しても同様のスペクトル整形を行うことによって、スペクトル整形後の最終的な合成音声のレベルで、自然音声に対する歪が小さくなるような代表音声素片を生成できるため、「めりはり」に優れたより明瞭な合成音声が得られる。

【００２２】本発明に係るさらに別の音声符号化方法は、複数の第１の音声素片のピッチおよび継続時間長の少なくとも一方に従って代表音声素片を用いて複数の合成音声素片を生成し、これらの合成音声素片と複数の第１の音声素片との間で定義される歪みの評価関数に基づいて複数の代表音声素片を求めて記憶し、これらの代表音声素片から所定の代表音声素片を選択して接続することによって音声を合成することを特徴とする。

【００２３】このように第２に音声素片から代表音声素片を選択するのでなく、第２の音声素片を用いずに第１の音声素片に対して最適な代表音声素片を生成することも可能である。このようにして生成された代表音声素片から、先と同様に所定の代表音声素片を選択して接続することにより、自然音声に近い高品質の合成音声が生成される。

【００２４】

【発明の実施の形態】以下、図面を参照して本発明の一実施形態を説明する。図１は、本発明の一実施形態に係る音声合成方法を実現する音声合成装置の構成を示すブロック図であり、大きく分けて合成単位学習系１と規則合成系２からなる。実際にテキスト音声合成を行う場合に動作するのは規則合成系２であり、合成単位学習系１は事前に学習を行って代表音声素片を生成するものである。

【００２５】まず、合成単位学習系１について説明する。合成単位学習系１は、代表音声素片とこれに付随する音素環境クラスタを生成する代表音声素片生成部１１と代表音声素片記憶部１２および音素環境クラスタ記憶部１３により構成される。代表音声素片生成部１１には、第１の音声素片であるトレーニング音声素片１０１とこれにラベル付けされた音素環境１０２および第２の音声素片である入力音声素片１０３が入力される。

【００２６】代表音声素片生成部１１では、トレーニング音声素片１０１にラベル付けされた音素環境１０２に含まれるピッチ周期および継続時間長の情報に従って、入力音声素片１０３のピッチ周期および継続時間長を変更することで複数の合成音声素片が内部的に生成され、さらにこれらの合成音声素片とトレーニング音声素片１０１との距離尺度に従って、代表音声素片１０４と音素環境クラスタ１０５が生成される。音素環境クラスタ１０５は、トレーニング音声素片１０１を後述するように

音素環境に関するクラスタに分類して生成される。

【００２７】代表音声素片１０４は代表音声素片記憶部１２に記憶され、音素環境クラスタ１０５は代表音声素片１０４と対応付けられて音素環境クラスタ記憶部１３に記憶される。代表音声素片生成部１１の処理については、後に詳細に説明する。

【００２８】次に、規則合成系２について説明する。規則合成系２は、代表音声素片記憶部１２と音素環境クラスタ記憶部１３と素片選択部１４および音声合成部１５により構成され、代表音声素片記憶部１２と音素環境クラスタ記憶部１３を合成単位学習系１と共有している。

【００２９】素片選択部１４には、入力音素の情報として、例えばテキスト音声合成のために入力テキストの形態素解析・構文解析後さらにアクセントやイントネーション処理を行って得られた韻律情報１１１と音韻記号列１１２が入力される。韻律情報１１１には、ピッチパターンおよび音韻継続時間長が含まれている。素片選択部１４では、これらの韻律情報１１１と音韻記号列１１２から入力音素の音素環境を内部的に生成する。

【００３０】そして、素片選択部１４は音素環境クラスタ記憶部１３より読み出された音素環境クラスタ１０６を参照して、入力音素の音素環境がどの音素環境クラスタに属するかを探索し、探索した音素環境クラスタに対応する代表音声素片選択情報１０７を代表音声素片記憶部１２へ出力する。

【００３１】音声合成部１５は、代表音声素片選択情報１０７に従って代表音声素片記憶部１２より選択的に読み出された代表音声素片１０８に対して、韻律情報１１１に従ってピッチ周期および音韻継続時間長を変更するとともに、素片の接続を行って合成音声信号１１３を出力する。ここで、ピッチおよび継続時間長を変更して素片を接続し音声を合成する方法としては、例えば残差駆動ＬＳＰ方法や波形編集方法など公知の技術を用いることができる。

【００３２】次に、本発明の特徴をなす代表音声素片生成部１１の処理の実施形態について具体的に説明する。図２のフローチャートは、代表音声素片生成部１１の第１の実施形態による処理手順を示している。

【００３３】この第１の実施形態による代表音声素片生成処理では、まず準備段階として連続発声された多数の音声データに対して音韻毎にラベリングを行い、ＣＶ，ＶＣＶ，ＣＶＣなどの合成単位に従って、トレーニング音声素片T_i（ｉ＝１，２，３，…，N_T）を切り出す。また、各トレーニング音声素片T_iに対応する音素環境P_i（ｉ＝１，２，３，…，N_T）も抽出しておく。ただし、N_Tはトレーニング音声素片の個数を表す。音素環境P_iは、少なくともトレーニング音声素片T_iの音韻とそのピッチおよび継続時間長の情報を含むものとし、その他に必要に応じて前後の音素などの情報を含むものとする。

9

【００３４】次に、上述したトレーニング音声素片T_iの作成と同様の方法により、多数の入力音声素片S_j（$i=1, 2, 3, \cdots, N_s$）を作成する。ただし、N_sは入力音声素片の個数を表す。ここで、入力音声素片S_jとしてはトレーニング音声素片T_iと同じものを使用してもよいし（すなわち$T_i = S_i$）、トレーニング音声素片T_iとは異なる音声素片を作成してもよい。いずれにしても、豊富な音韻環境を有する多数のトレーニング音声素片および入力音声素片が用意されていることが望ましい。

【００３５】このような準備段階を経た後、まず音声合成ステップＳ２１で、音素環境P_iに含まれるピッチおよび継続時間長に等しくなるように、入力音声素片S_jのピッチおよび継続時間長を変更して音声を合成することにより、合成音声素片G_{ij}を生成する。ここでのピッチおよび継続時間長の変更は、音声合成部１５におけるピッチおよび継続時間長の変更と同様の方法で行われるものとする。全ての音素環境P_i（$i=1, 2, 3, \cdots, N_T$）に従って入力音声素片S_i（$j=1, 2, 3, \cdots, N_s$）を用いて音声の合成を行うことにより、$N_T \times N_s$個の合成音声素片G_{ij}（$i=1, 2, 3, \cdots, N_T$、$j=1, 2, 3, \cdots, N_s$）を生成する。

【００３６】次に、歪み評価ステップＳ２２では、合成音声素片G_{ij}の歪みe_{ij}の評価を行う。この歪みe_{ij}の評価は、合成音声素片G_{ij}とトレーニング音声素片T_iとの間の距離尺度を求めることにより行う。距離尺度には、何らかのスペクトル距離を用いることができる。例えば、合成音声素片G_{ij}およびトレーニング音声素片T_iについて、ＦＦＴ（高速フーリエ変換）などを用いてパワースペクトルを求めて各パワースペクトル間の距離を評価する方法や、あるいは線形予測分析を行ってＬＰＣまたはＬＳＰパラメータなどを求めて各パラメータ間の距離を評価する方法などがある。その他にも、短時間フーリエ変換やウェーブレット変換などの変換係数を用いて評価する方法も用いることができる。また、各素片のパワーを正規化した上で歪みの評価を行う方法でもよい。

10

【００３７】次に、代表音声素片生成ステップＳ２３では、ステップＳ２２で得られた歪みe_{ij}に基づいて、入力音声素片S_jの中から指定された代表音声素片数Ｎの代表音声素片D_k（$k=1, 2, 3, \cdots, N$）を選択する。

【００３８】代表音声素片選択法の一例を説明する。入力音声素片S_jの中から選択されたＮ個の音声素片の集合$U = \{u_k \mid u_k = S_j\ (k=1, 2, 3, \cdots, N)\}$に対して、歪みの総和を表す評価関数$E_{D1}(U)$を次式（１）のように定義する。

【００３９】

【数１】

$$E_{D1}(U) = \sum_{i=1}^{N_T} \min(e_{ij1}, e_{ij2}, e_{ij3}, \cdots, e_{ijN}) \quad (1)$$

【００４０】ただし、$\min(e_{ij1}, e_{ij2}, e_{ij3}, \cdots, e_{ijN})$は$e_{ij1}, e_{ij2}, e_{ij3}, \cdots, e_{ijN}$の中の最小値を表す関数である。集合Ｕの組合せは$N_s ! / \{N ! (N_s - N) !\}$通りあり、これらの音声素片の集合Ｕの中から評価関数$E_{D1}(U)$を最小にするＵを探索し、その要素u_kを代表音声素片D_kとする。

【００４１】最後に、音素環境クラスタ生成ステップＳ２４では、音素環境P_i、歪みe_{ij}および代表音声素片D_kより、音素環境に関する複数のクラスタ（音素環境クラスタ）C_k（$k=1, 2, 3, \cdots, N$）を生成する。音素環境クラスタC_kは、例えば次式（２）で表されるクラスタリングの評価関数E_{C1}を最小化するクラスタを探索することによって得られる。

【００４２】

【数２】

$$E_{C1} = \sum_{k=1}^{N} \sum_{P_i \in C_k} e_{ijk} \quad (2)$$

【００４３】こうしてステップＳ２３およびＳ２４で生成された代表音声素片D_kおよび音素環境クラスタC_kは、図１の代表音声素片記憶部１２および音素環境クラスタ記憶部１３にそれぞれ記憶される。

【００４４】次に、図３のフローチャートを参照して代表音声素片生成部１１の第２の実施形態による処理手順について説明する。この第２の実施形態による代表音声素片生成処理では、まず初期音素環境クラスタ生成ステップＳ３０において、何らかの先見的な知識に基づいて予め音素環境のクラスタリングを行い、初期音素環境クラスタを生成する。音素環境のクラスタリングには、例えば音韻によるクラスタリングを行うことができる。

【００４５】そして、入力音声素片S_jおよびトレーニング音声素片T_iのうち音韻が一致する音声素片のみをそれぞれ用いて、図２のステップＳ２１，Ｓ２２，Ｓ２３，Ｓ２４と同様の合成音声素片生成ステップＳ３１、歪み評価ステップＳ３２、代表音声素片生成ステップＳ３３、音素環境クラスタ生成ステップＳ３４の処理を順次行い、全ての初期音素環境クラスタについて同様の操作を繰り返すことにより、全ての代表音声素片およびそれに対応する音素環境クラスタの生成を行う。こうして生成された代表音声素片および音素環境クラスタは、図

11

１の代表音声素片記憶部１２および音素環境クラスタ記憶部１３にそれぞれ記憶される。

【００４６】ただし、各初期音素環境クラスタ当たりの代表音声素片数が１であれば、初期音素環境クラスタが代表音声素片の音素環境クラスタとなるため、音素環境クラスタ生成ステップＳ３４は不要となり、初期音素環境クラスタを音素環境クラスタ記憶部１３に記憶すればよい。

【００４７】次に、図４のフローチャートを参照して代表音声素片生成部１１の第３の実施形態による処理手順を説明する。この第３の実施形態による代表音声素片生＊

12

＊成処理では、図２に示した第１の実施形態と同様に音声合成ステップＳ４１および歪み評価ステップＳ４２を順次経た後、次の音素環境クラスタ生成ステップＳ４３において、音素環境P_i および歪みe_{ij}に基づいて音素環境に関するクラスタC_k （ｋ＝１，２，３，…，Ｎ）を生成する。音素環境クラスタC_k は、例えば次式（３）（４）で表わされるクラスタリングの評価関数E_{c2}を最小化するクラスタを探索することによって得られる。

【００４８】

【数３】

$$E_{c2}=\sum_{k=1}^{N}\min\{f(k,1),f(k,2),f(k,3),\cdots,f(k,N)\} \quad (3)$$

$$f(k,j)=\sum_{P_i\in C_k}e_{ij} \quad (4)$$

【００４９】次に、代表音声素片生成ステップＳ４４において、歪みe_{ij}に基づいて音素環境クラスタC_k のそれぞれに対応する代表音声素片D_k を入力音声素片S_j より選択する。この代表音声素片D_k は、入力音声素片S_j から例えば次式（５）で表される歪み評価関数E_{D2}※

※(j) を最小化する音声素片を探索することによって得られる。

【００５０】

【数４】

$$E_{D2}(j)=\sum_{P_i\in C_k}e_{ij} \quad (5)$$

【００５１】なお、この第３の実施形態による代表音声素片生成処理を変形し、第２の実施形態と同様に、何らかの先見的な知識に基づいて予め生成した初期音素環境クラスタ毎に代表音声素片の生成および音素環境クラスタの生成を行うことも可能である。

【００５２】次に、図５～図９を用いて本発明の他の実施形態について説明する。図５は、本発明の他の実施形態に係る音声合成方法を実現する音声合成装置の構成を示すブロック図である。図１と相対応する部分に同一の参照符号を付して相違点を中心に説明すると、本実施形態では音声合成部１５の後段に適応ポストフィルタ１６が追加されている点が先の実施形態と異なり、これに加えて代表音声素片生成部１１における複数の合成音声素片の生成法も先の実施形態と異なっている。

【００５３】すなわち、代表音声素片生成部１１では先の実施形態と同様に、トレーニング音声素片１０１にラベル付けされた音素環境１０２に含まれるピッチ周期および継続時間長の情報に従って、入力音声素片１０３のピッチ周期および継続時間長を変更することで複数の合成音声素片を内部的に生成した後、これらの合成音声素片に対して適応ポストフィルタによるフィルタリングを施してスペクトル整形を行う。そして、この適応ポストフィルタによりスペクトル整形を行った後の各合成音声素片とトレーニング音声素片１０１との距離尺度に従って、代表音声素片１０４と音素環境クラスタ１０５が生成される。音素環境クラスタ１０５は、先の実施形態と同様にトレーニング音声素片１０１を音素環境に関するクラスタに分類して生成される。

【００５４】なお、この代表音声素片生成部１１において音素環境１０２に含まれるピッチ周期および継続時間長の情報に従って入力音声素片１０３のピッチ周期および継続時間長を変更して生成される複数の合成音声素片に対してフィルタリングを施してスペクトル整形を行う適応ポストフィルタは、音声合成部１５の後段に配置される適応ポストフィルタ１６と同様の構成でよい。

【００５５】一方、音声合成部１５では先の実施形態と同様に代表音声素片選択情報１０７に従って代表音声素片記憶部１２より選択的に読み出された代表音声素片１０８に対し、韻律情報１１１に従ってピッチ周期および音韻継続時間長を変更するとともに、素片の接続を行って合成音声信号１１３を生成するが、本実施形態ではこの合成音声信号１１３がさらに適応ポストフィルタ１６に入力され、ここで音質向上のためのスペクトル整形が行われた後、最終的な合成音声信号１１４が取り出される。

【００５６】図６に、適応ポストフィルタ１６の一構成例を示す。この適応ポストフィルタ１６は、ホルマント強調フィルタ２１とピッチ強調フィルタ２２を縦続配置して構成される。

【００５７】ホルマント強調フィルタ２１は、代表音声素片選択情報１０７に従って代表音声素片記憶部１２から選択的に読み出された代表音声素片１０８をＬＰＣ分析して得られるＬＰＣ係数に基づいて決定されるフィルタ係数に従って、音声合成部１５から入力される合成音

声信号１１３をフィルタリングすることにより、スペクトルの山の部分を強調する処理を行う。一方、ピッチ強調フィルタ２２は、韻律情報１１１に含まれるピッチ周期に基づいて決定されるパラメータに従って、ホルマント強調フィルタ２１の出力をフィルタリングすることにより、音声信号のピッチを強調する処理を行う。なお、ホルマント強調フィルタ２１とピッチ強調フィルタ２２の配置順序は逆であってもよい。

【００５８】このような適応ポストフィルタ１６の適用によりスペクトルが整形され、「めりはり」のある明瞭な音声を再生可能な合成音声信号１１４が得られる。適応ポストフィルタ１６としては図６に示した構成のものに限られず、音声符号化や音声合成の分野で用いられる公知の技術に基づく種々の構成を採用することが可能である。

【００５９】このように本実施形態では、規則合成系２において音声合成部１５の後段に適応ポストフィルタ１６が配置される点を考慮して、合成単位学習系１においても代表音声素片生成部１１で音素環境１０２に含まれるピッチ周期および継続時間長の情報に従って入力音声素片１０３のピッチ周期および継続時間長を変更して生成される複数の合成音声素片に対し、同様に適応ポストフィルタによるフィルタリングを行っている。従って、適応ポストフィルタ１６を通した後の最終的な合成音声信号１１４と同様のレベルで、自然音声に対する歪みが小さくなるような代表音声素片を代表音声素片生成部１１において生成できるため、さらに自然音声に近い合成音声を生成することが可能となる。

【００６０】次に、図５における代表音声素片生成部１１の処理の実施形態について具体的に説明する。図７、図８および図９のフローチャートは、図５における代表音声素片生成部１１の第１、第２および第３の実施形態による処理手順を示している。図７、図８および図９では、先に説明した図２、図３および図４に示した処理手順における音声合成ステップＳ２１、Ｓ３１およびＳ４１の後に、ポストフィルタリングステップＳ２５、Ｓ３６およびＳ４５が追加されている。

【００６１】ポストフィルタリングステップＳ２５、Ｓ３６およびＳ４５では、前述した適応ポストフィルタによるフィルタリングを行う。すなわち、音声合成ステップＳ２１、Ｓ３１およびＳ４１で生成された合成音声素片G_{ij}に対し、入力音声素片S_iをＬＰＣ分析して得られるＬＰＣ係数に基づいて決定されるフィルタ係数に従ってフィルタリングを行うことにより、スペクトルの山の部分を強調するホルマント強調を行う。また、このホルマント強調後の合成音声素片に対し、さらにトレーニング音声素片T_iのピッチ周期に基づいて決定されるパラメータに従ってフィルタリングを行うことにより、ピッチ強調を行う。

【００６２】このようにして、ポストフィルタリングス

テップＳ２５、Ｓ３６およびＳ４５において、スペクトル整形を行う。このポストフィルタリングステップＳ２５、Ｓ３６およびＳ４５は、前述したように規則合成系２において音声合成部１５の後段に設けられる適応ポストフィルタ１６により合成音声信号１１３のスペクトル整形を行って音質の向上を図るポストフィルタリングを行うことを前提に、合成単位の学習を可能とする処理であり、この処理を適応ポストフィルタ１６による処理と組み合わせることによって、最終的に「めりはり」のある明瞭な合成音声信号１１４が生成される。

【００６３】次に、図１０～図１３を用いて本発明の別の実施形態について説明する。図１０は、本発明の他の実施形態に係る音声合成方法を実現する音声合成装置の構成を示すブロック図である。図１と相対応する部分に同一の参照符号を付して相違点を中心に説明すると、本実施形態では代表音声素片生成部３１に入力音声素片１０３が入力されていない点がこれまでの実施形態と異なっている。

【００６４】すなわち、本実施形態では先の実施形態のように入力音声素片１０３の中から選択した音声素片を代表音声素片１０４とするのではなく、トレーニング音声素片１０１に対して最適な代表音声素片１０４を計算によって新たに生成する。音素環境クラスタ１０５は、先の実施形態と同様にトレーニング音声素片１０１を音素環境に関するクラスタに分類して生成される。

【００６５】次に、図１０における代表音声素片生成部３１の処理の実施形態について具体的に説明する。図１１のフローチャートは、代表音声素片生成部３１の第１の実施形態による処理手順を示している。この第１の実施形態による代表音声素片生成処理では、先の実施形態の代表音声素片生成部１１における代表音声素片生成処理と同様に、まず、準備段階として連続発声された多数の音声データに音韻毎にラベリングを行い、ＣＶ，ＶＣＶ，ＣＶＣなどの合成単位に従ってトレーニング音声素片T_i（$i=1, 2, 3, \cdots, N_T$）を切り出す。また、各トレーニング音声素片に対応する音素環境Ｐｉ（１，２，３，…，Ｎｒ）を抽出しておく。ただし、N_Tはトレーニング音声素片の個数を表す。音素環境は、少なくとも当該トレーニング音声素片の音韻とそのピッチパターンおよび継続時間長を含むものとし、その他に必要に応じて前後の音素などを含むものとする。

【００６６】このような準備段階を経た後、まず代表音声素片初期化ステップＳ５１で、指定された代表音声素片数Ｎの代表音声素片D_k（$k=1, 2, 3, \cdots, N$）を初期化して初期代表音声素片D_k^0（$k=1, 2, 3, \cdots, N$）を生成する。初期代表音声素片D_k^0としては、任意の音声素片を用いることが可能であり、例えばトレーニング音声素片T_iからランダムに選択された素片を用いることができる。

【００６７】次に、音声合成ステップＳ５２で、D_k^0

15

のピッチおよび継続時間長をP_iのピッチパターンおよび継続時間長に等しくなるように変更して音声を合成して合成音声素片G_{ik}を生成する。ここでのピッチおよび継続時間長の変更は、音声合成部20におけるピッチおよび継続時間長の変更と同様の方法で行われるものとする。全てのP_i（i＝1，2，3，…，N_T）に従ってD_k^0（k＝1，2，3，…，N_T）を用いて合成を行うことにより、N_T×N個の合成音声素片G_{ik}（i＝1，2，3，…，N_T、k＝1，2，3，…N）を生成する。

【0068】次に、歪み評価ステップS53では、合成音声素片G_{ik}とトレーニングとの間で定義される歪みe_{ij}の評価を行う。歪みの評価法としては、波形の2乗誤差や何らかのスペクトル距離を用いることができる。例えば、FFTなどを用いてパワースペクトルを求めてそ*

16

*の間の距離を求める方法や、あるいは線形予測分析を行ってLPCまたはLSPパラメータなどを求めてパラメータ間の距離を評価する方法などがある。その他にも、短時間フーリエ変換やウェーブレット変換などの変換係数を用いて評価する方法が考えられる。また、各素片のパワーを正規化した上で歪みの評価を行うことも考えられる。

【0069】次に、音素環境クラスタ生成ステップS54では音素環境P_iおよび歪みe_{ik}に基づいて、音素環境に関するクラスタC_k（k＝1，2，3，…，N）を生成する。音素環境クラスタC_kは、例えば次式で表されるクラスタリングの評価関数E_{c3}を最小化するクラスタを探索することによって得られる。

【0070】

【数5】

$$E_{c3}=\sum_{k=1}^{N}\sum_{P_i\in C_k} e_{ik} \qquad (6)$$

【0071】但し、次式に示されるように全ての音素環境クラスタC_k（k＝1，2，3，…，N）の和集合は音素環境の全体集合Aと等しく、かつ任意の異なる2つの音素環境クラスタの積集合は空集合ϕになるものとす※

※る。

【0072】

【数6】

$$C_1\cup C_2\cup\cdots\cup C_N=A \qquad (7)$$

$$C_i\cap C_j=\phi\ (i\neq j) \qquad (8)$$

【0073】次に、代表音声素片生成ステップS55では、代表音声素片を更新するため、クラスタC_kに対応する代表音声素片D_k^1をクラスタ毎に求める。代表音声素片D_k^1は、クラスタに属するトレーニング音声素片と、対応する合成音声素片との歪みの総和を表す評価★

★関数が最小になるように計算される。評価関数としては、例えば次式に示す波形の2乗誤差の総和を用いることができる。

【0074】

【数7】

$$ED_3(D_k)=\sum_{P_i\in C_k}|t_i-g_{ik}(D_k)|^2 \qquad (9)$$

【0075】但し、t_iはトレーニング音声素片T_iの波形を表すベクトル、$g_{ik}(D_k)$はP_iに従ってD_kを用いて合成された合成音声素片G_{ik}の波形を表すベクトルである。また、評価関数の例としては、これ以外にもパワースペクトルの距離やLPC，LSPなどのパラメータ間の距離の総和などがある。評価関数を代表音声素片で偏微分したものを0とおいた方程式が解ける場合には、これを解いて評価関数を最小にする代表音声素片を解析的に求めることができる。それ以外の場合は、公知の最適化手法を用いて代表音声素片を求めることができる。

【0076】このようにステップS52からS55までの処理で、初期代表音声素片D_k^0から代表音声素片D_k^1に更新される。代表音声素片が更新されたことによって音素環境クラスタが変化するため、代表音声素片および音素環境クラスタの変化が十分小さくなり収束するまで、ステップS52からS55までの処理を繰り返す必要がある。

【0077】そこで、次の収束判定ステップS56では、更新の前後における代表音声素片の変化の度合から代表音声素片および音素環境クラスタの変化が収束したか否かを判定し、収束していないと判定された場合はステップS52からS55までの処理を繰り返してさらに代表音声素片を更新し、収束したと判定された場合は処理を終了し、最新の代表音声素片D_k^m（k＝1，2，3，…，N、mは繰り返し回数）が代表音声素片D_k（k＝1，2，3，…，N）となる。

【0078】こうして生成された代表音声素片D_kおよび音素環境クラスタC_kは、図10の代表音声素片記憶部12および音素環境クラスタ記憶部13にそれぞれ記憶される。

【0079】次に、図12のフローチャートを参照して代表音声素片生成部31の第2の実施形態による処理手順について説明する。この第2の実施形態による代表音声素片生成処理では、まず初期音素環境クラスタ生成ステップS61において、何らかの先見的な知識に基づい

て予め音素環境のクラスタリングを行い、初期音素環境クラスタを生成する。音素環境のクラスタリングには、例えば音韻によるクラスタリングを行うことができる。

【００８０】そして、トレーニング音声素片T_iのうち音韻が一致する音声素片のみを用いて、図１１のステップＳ５１、Ｓ５２、Ｓ５３、Ｓ５４、Ｓ５５、Ｓ５６と同様の代表音声素片初期化ステップＳ６２、音声合成ステップＳ６３、歪み評価ステップＳ６４、音素環境クラスタ生成ステップＳ６５、代表音声素片生成ステップＳ６６、収束判定ステップＳ６７の処理を順次行い、全ての初期音素環境クラスタについて同様の操作を繰り返すことにより、全ての代表音声素片およびそれに対応する音素環境クラスタの生成を行う。こうして生成された代表音声素片および音素環境クラスタは、図１０の代表音声素片記憶部１２および音素環境クラスタ記憶部１３にそれぞれ記憶される。

【００８１】但し、各初期音素環境クラスタ当りの代表音声素片数が１であれば、初期音素環境クラスタが代表音声素片の音素環境クラスタとなるため、ステップＳ６２、Ｓ６３、Ｓ６４、Ｓ６５、Ｓ６７の処理は不要とな

＊り、初期音素環境クラスタに対応する代表音声素片を代表音声素片生成ステップＳ６６で求めればよい。この場合には、初期音素環境クラスタを音素環境クラスタ記憶部１３に記憶すればよい。

【００８２】次に、図１３のフローチャートを参照して代表音声素片生成部３１の第３の実施形態による処理手順について説明する。まず、図１１のステップＳ５１、Ｓ５２、Ｓ５３と同様の代表音声素片初期化ステップＳ７１、音声合成ステップＳ７１、歪み評価ステップＳ７３の処理を順次行って、合成音声素片G_{ik}とトレーニング音声素片T_iの間の歪みe_{ik}を求める。

【００８３】次に、トレーニング音声素片クラスタ生成ステップＳ７４では、歪みe_{ik}に基づいてトレーニング音声素片T_iのクラスタC'_k（$k=1, 2, 3, \cdots, N$）を生成する。このトレーニング音声素片クラスタC'_kは、例えば次式で表されるクラスタリングの評価関数E_{c4}を最小化するクラスタを探索することによって得られる。

【００８４】

【数８】

$$E_{c4}=\sum_{k=1}^{N}\sum_{T_i\in C'_k} c_{ik} \qquad (10)$$

【００８５】次に、図１１のステップＳ５５、Ｓ５６と同様の代表音声素片生成ステップＳ７５、収束判定ステップＳ７６の処理を順次行って、代表音声素片およびそれに対応するトレーニング音声素片クラスタが生成される。

【００８６】最後に、音素環境クラスタ生成ステップＳ７７では、トレーニング音声素片クラスタC'_kに属するトレーニング音声素片T_iに共通する音素環境を抽出して音素環境クラスタC_kを生成する。但し、音素環境クラスタC_k（$k=1, 2, 3, \cdots, N$）は、式（７）（８）の条件を満たすものとする。また、本実施形態の音声合成方法に前の実施形態と同様にポストフィルタリング処理を組み合わせることも可能である。

【００８７】

【発明の効果】以上説明したように、本発明の音声合成方法によれば、入力音声素片に対してピッチおよび継続時間長の少なくとも一方の変更を行って生成される合成音声のレベルで自然音声に対する歪みを評価し、その歪み評価結果に基づいて入力音声素片から選択した音声素片を代表音声素片とするか、あるいは歪み評価結果に基づいて代表音声素片を生成するため、音声合成装置の特性をも考慮した代表音声素片の生成が可能であり、この代表素片を接続して音声合成を行うことによって、自然音声に近い高品質の合成音声を生成することができる。

【００８８】また、本発明ではさらに代表音声素片の接続によって合成される音声に対してスペクトル整形を行うと共に、合成音声素片に対しても同様のスペクトル整形を行うことにより、スペクトル整形後の最終的な合成音声信号のレベルで、自然音声に対する歪が小さくなるような代表音声素片を生成できるため、「めりはり」のあるより明瞭な合成音声を生成することができる。

【図面の簡単な説明】

【図１】本発明の一実施形態に係る音声合成装置のブロック図

【図２】図１中の代表音声素片生成部での第１の実施形態による処理手順を示すフローチャート

【図３】図１の中の代表音声素片生成部での第２の実施形態による処理手順を示すフローチャート

【図４】図１中の代表音声素片生成部での第３の実施形態による処理手順を示すフローチャート

【図５】本発明の他の実施形態に係る音声合成装置のブロック図

【図６】図５中の適応ポストフィルタの構成例を示すブロック図

【図７】図５中の代表音声素片生成部での第１の実施形態による処理手順を示すフローチャート

【図８】図５の中の代表音声素片生成部での第２の実施形態による処理手順を示すフローチャート

【図９】図５中の代表音声素片生成部での第３の実施形態による処理手順を示すフローチャート

【図１０】本発明の別の実施形態に係る音声合成装置のブロック図

【図１１】図１０中の代表音声素片生成部での第１の実施形態による処理手順を示すフローチャート

19

【図１２】図１０の中の代表音声素片生成部での第２の実施形態による処理手順を示すフローチャート

【図１３】図１０中の代表音声素片生成部での第３の実施形態による処理手順を示すフローチャート

【符号の説明】

１…合成単位学習系

２…規則合成系

１１…代表音声素片生成部

１２…音素環境クラスタ記憶部

１３…代表音声素片記憶部

１４…素片選択部

１５…音声合成部

１６…適応ポストフィルタ

２１…ホルマント強調フィルタ

20

２２…ピッチ強調フィルタ

１０１…トレーニング音声素片（第１の音声素片）

１０２…トレーニング音声素片にラベル付けされた音素環境

１０３…入力音声素片（第２の音声素片）

１０４…代表音声素片

１０５…音素環境クラスタ

１０６…音素環境クラスタ

１０７…代表音声素片選択情報

１０８…代表音声素片

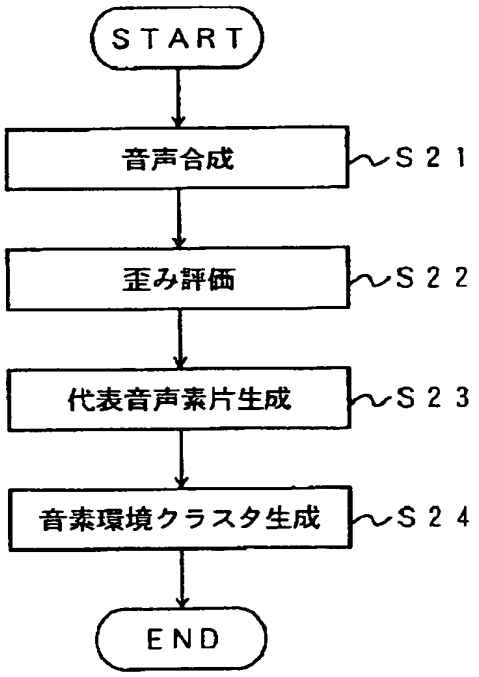
１１１…韻律情報

１１２…音韻記号列

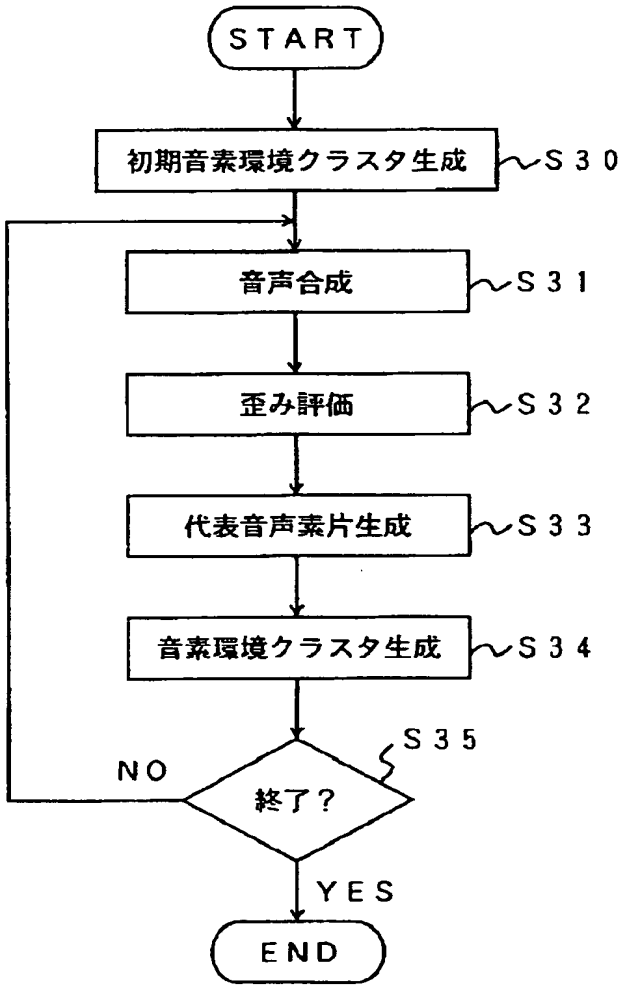
１１３…合成音声信号

１１４…合成音声信号

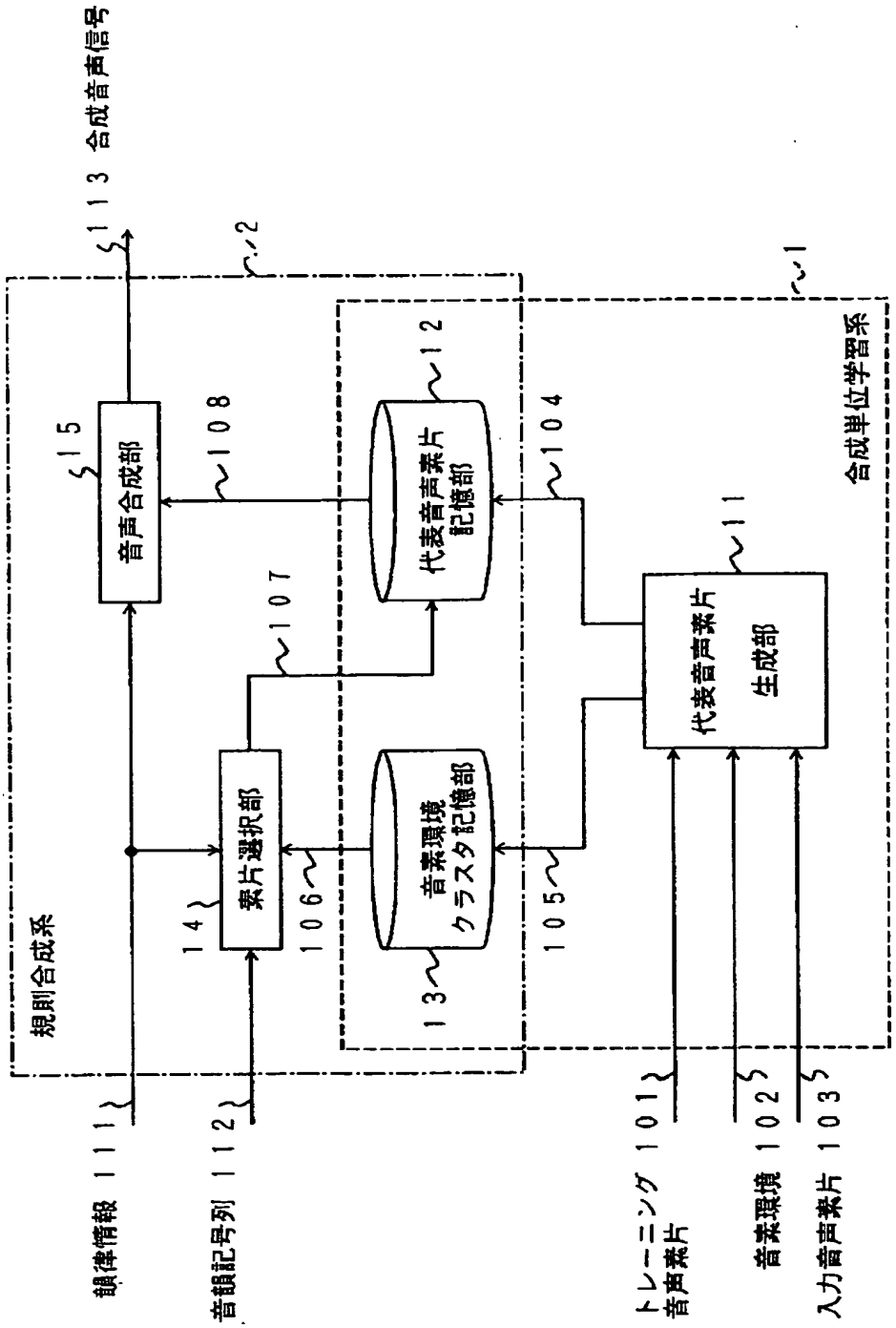
【図２】



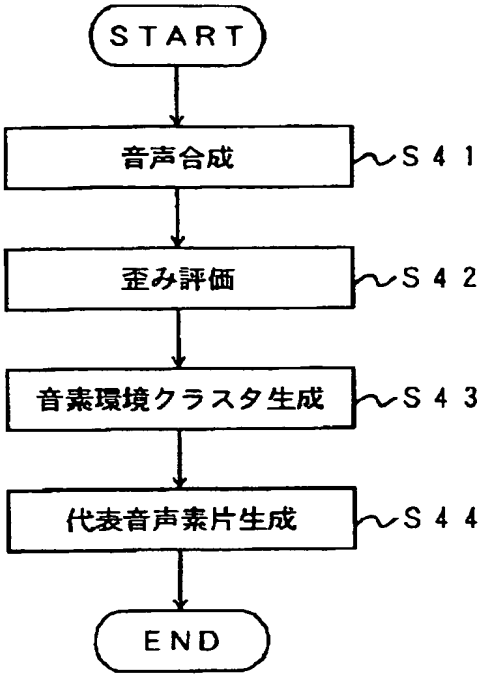
【図３】



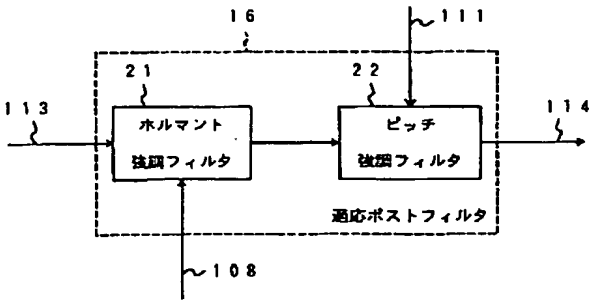
【図１】


１１３　合成音声信号
１５
音声合成部
１０８
１２
代表音声素片記憶部
１０４
１１
代表音声素片生成部
１０７
１４
素片選択部
１０６
音素環境クラスタ記憶部
１３
１０５
規則合成系
合成単位学習系
１
２
韻律情報　１１１
音韻記号列　１１２
トレーニング音声素片　１０１
音素環境　１０２
入力音声素片　１０３

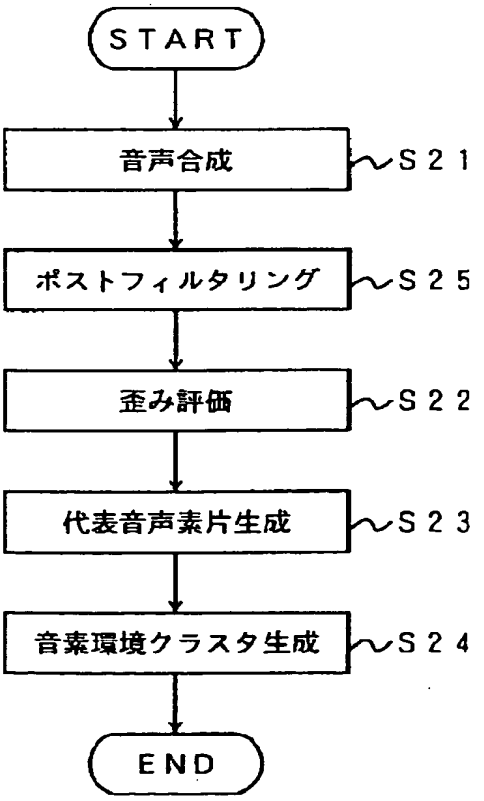
【図４】


START
音声合成
Ｓ４１
歪み評価
Ｓ４２
音素環境クラスタ生成
Ｓ４３
代表音声素片生成
Ｓ４４
END

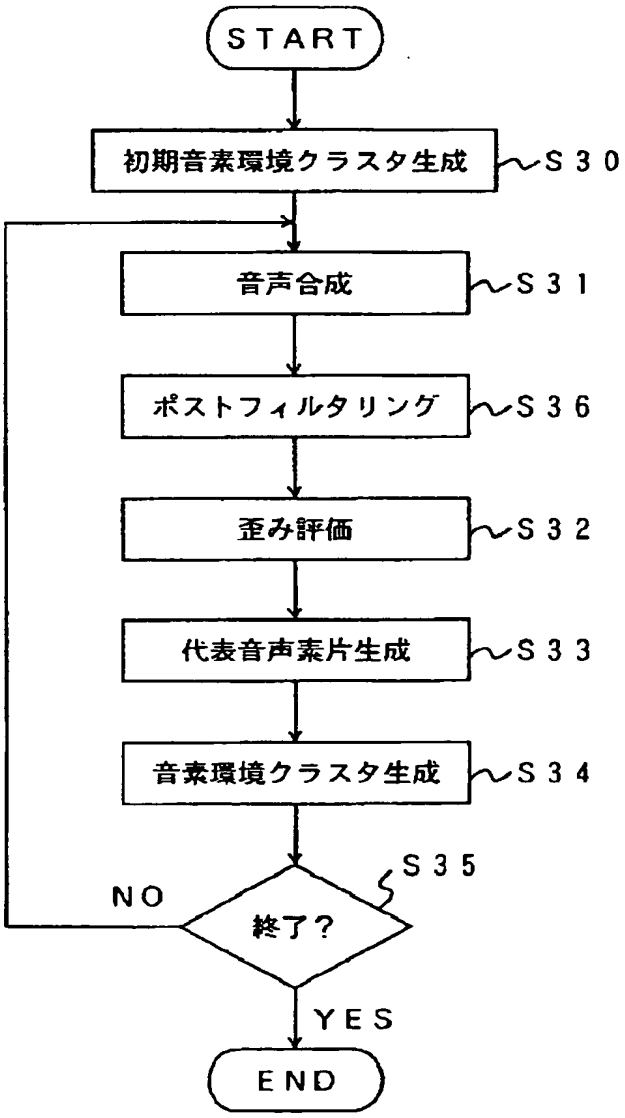
【図６】


１６
１１１
２１
２２
１１３
ホルマント
強調フィルタ
ピッチ
強調フィルタ
１１４
適応ポストフィルタ
１０８

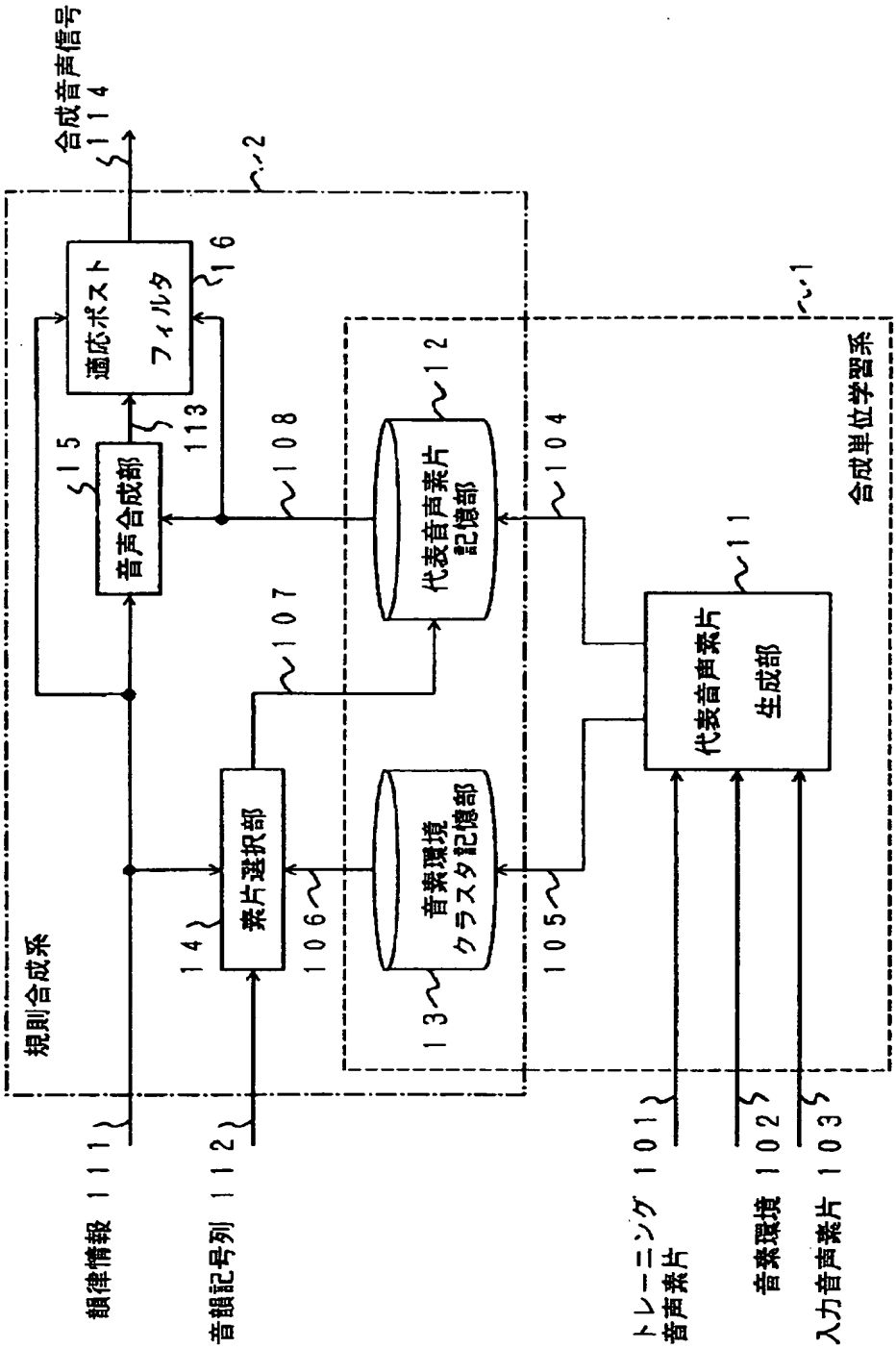
【図７】


START
音声合成
Ｓ２１
ポストフィルタリング
Ｓ２５
歪み評価
Ｓ２２
代表音声素片生成
Ｓ２３
音素環境クラスタ生成
Ｓ２４
END

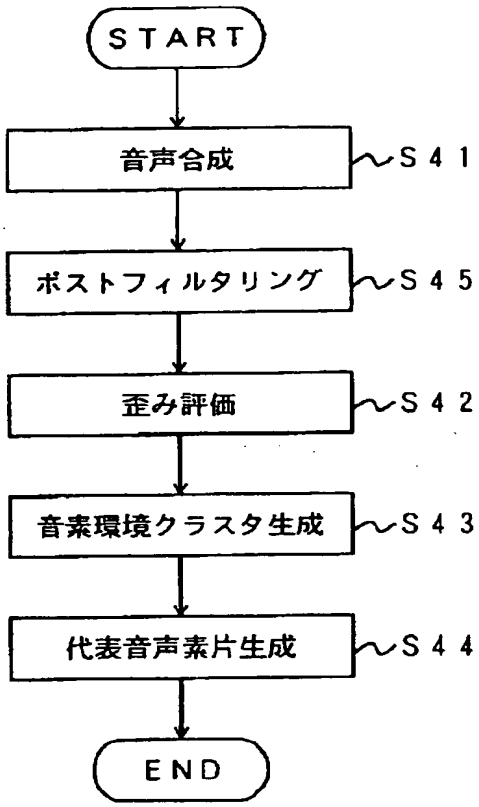
【図８】


START
初期音素環境クラスタ生成
Ｓ３０
音声合成
Ｓ３１
ポストフィルタリング
Ｓ３６
歪み評価
Ｓ３２
代表音声素片生成
Ｓ３３
音素環境クラスタ生成
Ｓ３４
Ｓ３５
NO
終了？
YES
END

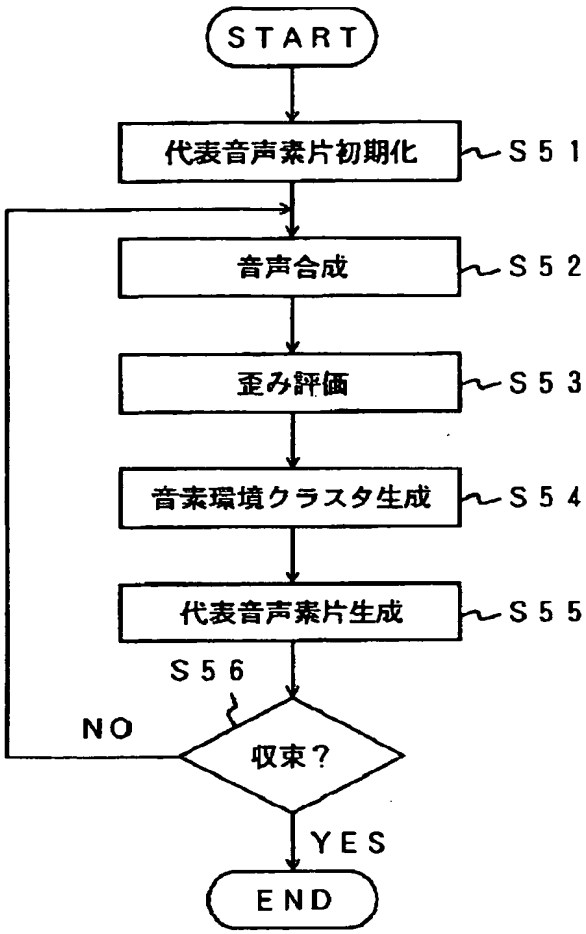
【図５】


合成音声信号
114
2
適応ポスト
フィルタ
16
15
音声合成部
113
108
代表音声素片
記憶部
12
104
1
合成単位学習系
107
代表音声素片
生成部
11
規則合成系
素片選択部
14
106
音素環境
クラスタ記憶部
13
105
韻律情報 111
音韻記号列 112
トレーニング
音声素片 101
音素環境 102
入力音声素片 103

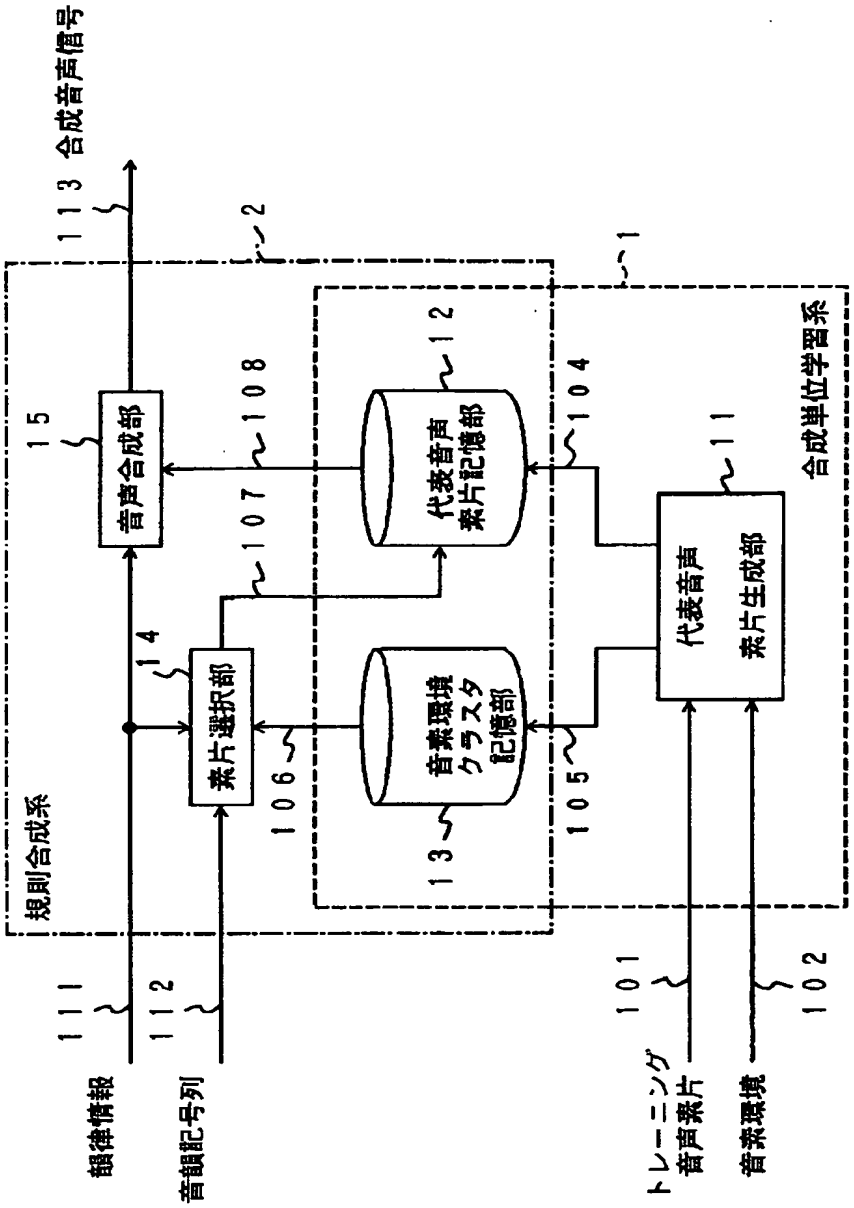
【図９】


START
音声合成
S41
ポストフィルタリング
S45
歪み評価
S42
音素環境クラスタ生成
S43
代表音声素片生成
S44
END

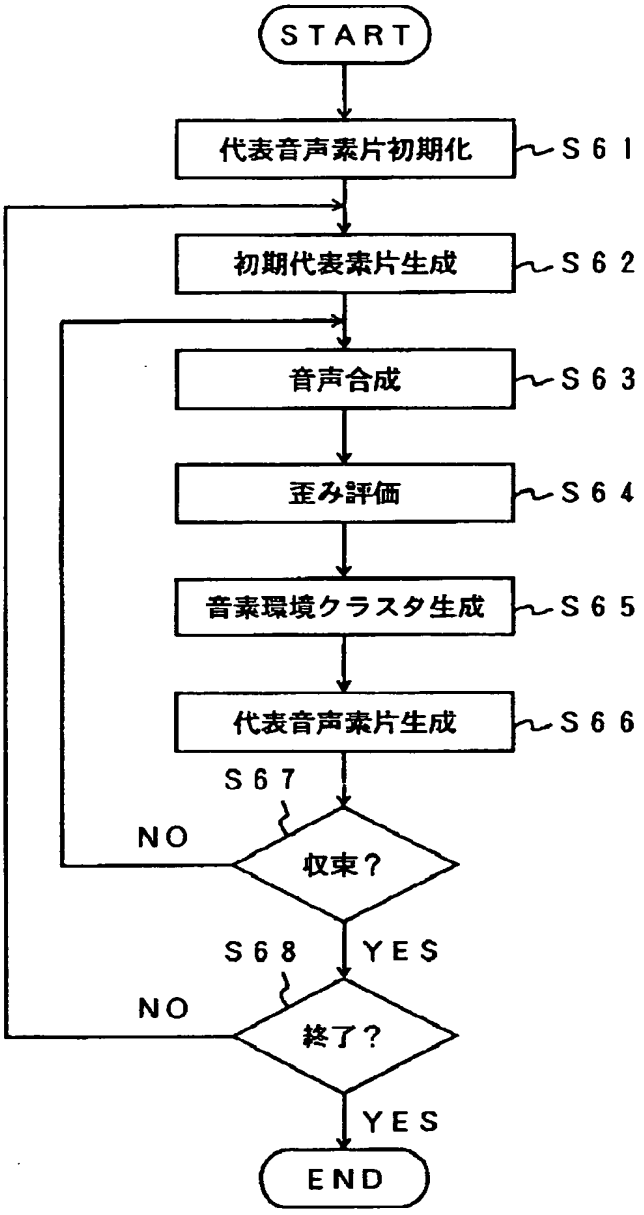
【図１１】


START
代表音声素片初期化
S51
音声合成
S52
歪み評価
S53
音素環境クラスタ生成
S54
代表音声素片生成
S55
S56
NO
収束？
YES
END

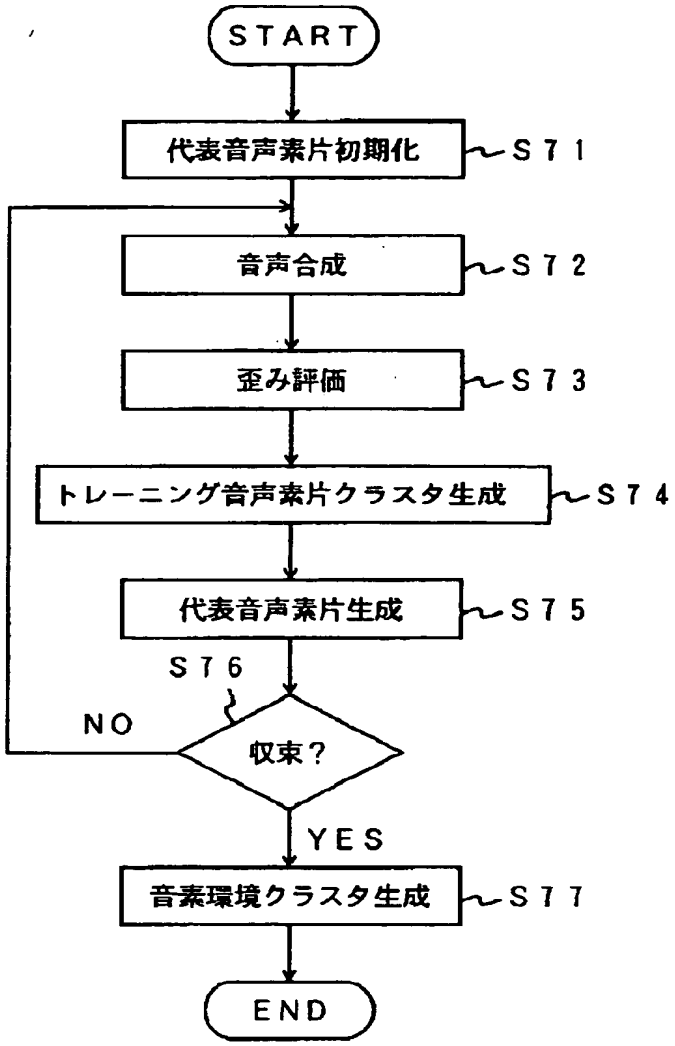
【図１０】


113 合成音声信号
2
1
規則合成系
合成単位学習系
15 音声合成部
108
12 代表音声素片記憶部
104
107
11 代表音声素片生成部
14 素片選択部
106
音素環境クラスタ記憶部
105
13
111 韻律情報
112 音韻記号列
101 トレーニング音声素片
102 音素環境

【図１２】


START
代表音声素片初期化
S61
初期代表素片生成
S62
音声合成
S63
歪み評価
S64
音素環境クラスタ生成
S65
代表音声素片生成
S66
S67
収束？
NO
YES
S68
終了？
NO
YES
END

【図１３】


START
代表音声素片初期化
S71
音声合成
S72
歪み評価
S73
トレーニング音声素片クラスタ生成
S74
代表音声素片生成
S75
S76
収束？
NO
YES
音素環境クラスタ生成
S77
END